014524-13-T
SEL 140

Evaluation of Computing Systems Using Functionals of a Stochastic Process

(NASA-CR-163512) EVALUATION OF COMPUTING SYSTEMS USING FUNCTIONALS OF A STOCHASTIC PROCESS (Michigan Univ.) 45 p HC A03/MF A01 CSCL 09B

N80-33102

Unclas
G3/60 28783

J. F. MEYER
L. T. WU

July 1980

Prepared for

National Aeronautics and Space Administration
Langley Research Center
Hampton, Virginia 23365

G.E. Migneault, Technical Officer

NASA Grant NSG 1306

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
SYSTEMS ENGINEERING LABORATORY
THE UNIVERSITY OF MICHIGAN, ANN ARBOR

014524-13-T
SEL 140

EVALUATION OF COMPUTING SYSTEMS USING FUNCTIONALS OF A STOCHASTIC PROCESS

J. F. Meyer
L. T. Wu

Prepared For
National Aeronautics and Space Administration
Langley Research Center
Hampton, Virginia 23365
G. E. Magneault, NASA Technical Officer
NASA Grant NSG 1306

John F. Meyer, Principal Investigator
Department of Electrical and Computer Engineering
Systems Engineering Laboratory
The University of Michigan
Ann Arbor, Michigan 48109

July 1980

TABLE OF CONTENTS

1. Introduction

When modeling degradable computing systems by stochastic processes for the purpose of performance and reliability evaluation, the models employed are typically Markov processes (see [1], for example) or models which can be analyzed in terms of imbedded Markov processes (semi-Markov processes, branching processes, etc.; see [2]). However, to ensure the validity of the Markov assumption, it is usually necessary to model the structure and behavior of the system at a low level, e.g., a level describing the system's structural resources (processing units, memory units, input buffers (queues), etc.). Performance and reliability measures, on the other hand, often quantify the system's behavior in terms of high level, user oriented variables (throughput, response time, operational status, etc.) which, if viewed as stochastic processes, are seldom Markovian. In such cases, an essential part of the modeling effort is to establish a "connection" between the low and high levels, so as to resolve the probabilistic nature of the measure in question.

Historically, in the context of reliability modeling, this connection has taken a form that lies at one of two extremes. At one extreme, system "success" is defined in terms of the underlying structural resources (at least so many fault-free processors, at least so many fault-free memory units, etc.), in which case the connection between structure (available fault-free resources) and performance (success or failure) is immediate. At the other extreme, the object of the modeling effort is the connection, per se, and the resulting model is typically some form of "event-tree" or "fault-free" (see [3], for example).

More recently, in efforts to deal simultaneously with issues of performance and reliability, the general nature of this connection has been formalized and is referred to as a "capability function" [4] of the system. In this setting, a "total system" S, comprising a computing system and its computational environment, is modeled at a low level by a stochastic process X_S (the "base model" of S). Then, relative to a high level variable Y_S (the "performance" of S), the capability function of S is a function γ_S which translates state trajectories (sample paths) of the process X_S into corresponding values of the performance variable Y_S. Knowing X_S and γ_S, it is possible (although not necessarily feasible) to solve for the probability distribution function of Y_S and, hence, determine the "performability" of S.

In cases where the performance Y_S is relatively far removed from the base model X_S, solution procedures can be facilitated by introducing intermediate models at levels between X_S and Y_S. One use of such a model hierarchy is a step-by-step formulation of the inverse of γ_S, beginning at Y_S and terminating at the base model X_S. If Y_S is discrete, the performability of S can then be solved by determining the probabilities of certain trajectory sets that correspond (under γ_S^{-1}) to performance values of Y_S. (See [4] for details regarding this method; also see [5] which uses this technique to evaluate the performability of the SIFT computer.)

Another role that can be played by an intermediate model, and the one we explore in this report, is to represent the probabilistic nature of S at a level which is higher than the base model and thus "closer" to the performance variable. In particular, we consider a

class of intermediate models which are generally referred to as "functionals" of a Markov process (see [6], for example). More precisely, a functional of a Markov process X is a stochastic process $\overline{X}$ (not necessarily Markovian) where, relative to a specified real-valued function f defined in the states of X, if $X_t=i$ then $\overline{X}_t=f(i)$ for each time instant t in the utilization period. We first discuss (Section 2) the conditions under which $\overline{X}$ is itself a Markov process, in which case $\overline{X}$ can be viewed as a "lumped" version of X. Since these conditions are quite severe, we go on to examine solution methods for non-Markovian functionals (Section 3) where, in particular, we develop a closed-form solution of performability for the case where performance is identified with the minimum value of a functional.

2. Functionals as Intermediate Models

When describing system behavior in user oriented terms, it is possible, in many cases, to identify various operational "modes" for the system (including a failure mode) which result in different degrees of user satisfaction. Moreover, for a given mode of operation, the extent of user satisfaction can often be quantified as a (real number) "rate" at which that operation benefits (or penalizes) the user. Depending on the application, these rates can have a variety of interpretations relating to the system's productivity, responsiveness, etc. or, at a higher level, to such things as economic benefit, e.g., the "worth rate" (measured, say, in dollars/unit time) associated with a given mode of operation.

Under the above conditions, a user-oriented model can be constructed in a rather natural fashion. As in our introductory remarks (and following the terminology and notation of [4]), let S denote the total system in question and suppose that we have already determined a base model X_S and a capability function γ_S (relative to some specified performance variable Y_S). Suppose further that the base model process X_S is defined relative to a continuous time interval T (the "utilization period"), that is,

$$X_S = \{X_t \mid t \in T\} \tag{1}$$

where the random variables X_t take values in a discrete (countable) state space Q. Thus, without loss of generality, the states of X_S will often be designated by positive integers, viz.

$$Q = \{1,2,3,\ldots\}$$

or, in case Q is finite,

$$Q = \{1,2,\ldots,n\}.$$

Finally, we presume that at the base level, the system model is Markovian with a time-invariant structure, that is, X_S is a (continuous-time) time-homogeneous Markov process. (Since Q is discrete, some authors would refer to X_S as a Markov "chain"; we choose to maintain the use of the generic term "process".)

Within this framework, let us now consider the situation discussed above where, at a higher level, one is able to identify various operational modes for S, each having an associated operational rate. If, further, each state of the base model can be classified according to some mode of operation, then there is a naturally defined real-valued function

$$f: Q \rightarrow \mathbb{R} \qquad (2)$$

where, for each $i \in Q$, $f(i)$ is the operational rate associated with the mode containing i. Moreover, if we let $\overline{Q}$ denote the range of f (i.e., $\overline{Q}=\{f(i)\,|\,i \in Q\}$) and, for each variable X_t of X_S (see (1)), we let

$$\overline{X}_t = f(X_t), \qquad (3)$$

it follows that

$$\overline{X}_S = \{\overline{X}_t \,|\, t \in T\} \qquad (4)$$

is a stochastic process with state space $\overline{Q}$ (referred to generally as a "functional" [6] of the underlying Markov process).

Except in the special case where f is 1-1 (i.e., different states correspond to diffferent modes of operation), the derived process $\overline{X}_S$ represents a simpler, higher level view of the system S. However, to qualify $\overline{X}_S$ as an "intermediate" model, we must also require that $\overline{X}_S$ be compatible with the performance variable Y_S to the extent that Y_S can be solved in terms of $\overline{X}_S$. More precisely, letting κ denote the

translation of trajectories of X_S to trajectories of $\overline{X}_S$ (i.e., $\kappa(u)=\overline{u}$ where $\overline{u}(t)=f(u(t))$, for all $t\varepsilon T$), there must exist a capability function $\overline{\gamma}_S$ for $\overline{X}_S$ such that

$$\overline{\gamma}_S \cdot \kappa = \gamma_S \qquad (5)$$

(where $\cdot$ denotes functional composition, first applying κ). Although the above condition appears somewhat formidable, it says simply that the higher level model $\overline{X}_S$ must remain detailed enough to permit solution of the system's performability. Moreover, this condition can be typically satisfied in practice if the definition of performance (i.e., Y_S) is taken into account when identifying the various modes of operation and assigning rates to these modes.

If f, as defined in (2), satisfies condition (5) then we refer to f as an operational structure of S and, since states inherit the rates assigned to modes, the value f(i) is referred to as the operational rate of i (or, when context permits, simply the "rate of i"). Likewise, the corresponding functional $\overline{X}_S$ is referred to as an operational model of S or, alternatively, a model of S at the operational level.

In reliability modeling where, at the operational level, a system is typically viewed as either "operating" or "not operating", the concept of an operational structure reduces to the familiar notion of a "structure function" [3]. Technically, a function $f: Q \to \mathbb{R}$ is a structure function if Q has binary coordinates, i.e., $Q=\{0,1\}^m$, and f(i) is 1 or 0 according as S is operating or not operating in state i . More recently, operational structures have been employed (at least implicitly) in the context of performance-reliability modeling where the operational rates are referred to as computational "capacities" [7],[8]. Although capacity (which typically

refers to the maximum rate at which a computer can "supply" computations) is a legitimate interpretation of "operational rate", it should be emphasized that, in general, such rates can represent an interaction of supply (by the computer) and demand (from the environment); this follows from the fact that, as generally conceived, a state i of the base model represents a particular status of both the computer and its environment; hence, both supply and demand can be accounted for when translating i, via f, to its corresponding operational rate f(i).

In various special forms, then, the concept of an operational structure is no stranger to performance and reliability modeling. On the other hand, the general nature of the associated "functional" $\overline{X}_S$, how it relates to the base model, how it can be exploited in solution procedures, etc., appear to be subjects that deserve further investigation.

To begin, let us suppose that S is modeled by (X_S, γ_S) and $f: Q \to \mathbb{R}$ is the operational structure of S. Then, to the extent that it is feasible, we would like to determine the probabilistic nature of the operational model, that is, the functional (see [6],[9])

$$\overline{X}_S = \{f(X_t) \mid t \in T\}$$

of the Markov process X_S. The first questions which arise in this regard are whether $\overline{X}_S$ is, itself, a Markov process and, if so, whether it is time-homogeneous. Note that, since these questions do not involve the actual values of f, $\overline{X}_S$ may be regarded here as a "lumped" [10] version of X_S, where states i and j are in the same lump if and only if f(i)=f(j). (In other words, lumps coincide with the operational "modes" of S.) Moreover, these questions are obviously independent of our interpretations of X_S and $\overline{X}_S$ and hence, in the

development that follows, we can drop the specific reference to S.

Let us suppose, then, that X is a time-homogeneous Markov process with state space Q and, given a function $f: Q \to \mathbb{R}$, let $\overline{X}$ denote the corresponding functional or, what is the same, the lumped process induced by f. If f is 1-1 then $\overline{X}$ is obviously both Markovian and time-homogeneous since, in this case, the lumping is trivial. If f is properly a many-to-one function the answers are no longer obvious and, indeed, the questions need further clarification.

In the latter regard, let us restrict our attention to finite-state processes X which are "regular" in the sense that their transition probabilities are uniquely determined by a "generator matrix" or, equivalently, a "state-transition-rate" diagram. More precisely, let X be a Markov process with state space $Q=\{1,2,\ldots,n\}$ and, for $i \neq j$, let λ_{ij} denote the transition rate from state i to state j. Then the generator matrix of X is the n×n matrix

$$A = [a_{ij}]$$

where

$$a_{ij} = \begin{cases} \lambda_{ij} & \text{if } i \neq j, \\ -\sum_{k \neq i} \lambda_{ik} & \text{if } i = j. \end{cases} \tag{6}$$

If, further, we let P(t) denote the "transition function" of X (we presume here that $t \in [0,\infty)$), i.e.,

$$P(t) = [p_{ij}(t)]$$

where

$$p_{ij}(t) = \Pr[X_t = j \mid X_0 = i],$$

then A uniquely determines P(t) by the well known equation

$$P(t) = e^{At}.$$

Hence A (or P(t)) together with an initial distribution

$$p = [p_1 \cdots p_n]; \; p_i = Pr[X_0=i]$$

are enough to completely specify the process X.

In the above setting, our original inquiry thus reduces to the question:

Q1) Given A, p and f, is $\overline{X}$ Markovian?

In many applications, however, one wants the freedom to alter the initial distribution p without losing the Markov property. In this case we are asking:

Q2) Given A and f, is $\overline{X}$ Markovian for arbitrary p?

Finally, we can raise our sights even higher and ask:

Q3) Given A and f, is $\overline{X}$ Markovian for arbitrary p and, moreover, is the transition function of $\overline{X}$ independent of p?

Adopting the terminology of [10] (which investigates the discrete-time versions of Q1 and Q3), if the answer to Q1 is "yes" then the process X_S specified by A and p is "weakly lumpable" with respect to f. A "yes" answer to Q2 is stronger but, generally, these Markov processes will not be time-homogeneous. If the answer to Q3 is "yes" then, for all initial distributions p, the Markov processes $\overline{X}$ have the same transition function and, by the homogeneity of X, it follows that this function is invariant under time shifts, i.e., the processes are time-homogeneous. In this case we say that the processes X specified by A are "strongly lumpable" with respect to f.

Addressing first the question of weak lumpability (Q1), if $\overline{X}$ is to be a Markov process, we must insure it has the "memoryless" property, that is, any sequence of past observations of $\overline{X}$ provides the same

information as the last of those observations. To formalize this requirement, if $0 \le t_1 < t_2 < \cdots < t_k$ is a sequence of observation times and $q_i \varepsilon \overline{Q} = f(Q)$ is the state of $\overline{X}$ observed at time t_i, for each underlying state $j \varepsilon \{1,\ldots,n\}$, let

$$M_j(t_1,\ldots,t_k;q_1,\ldots,q_k) = \Pr[X_{t_k}=j \mid \overline{X}_{t_1}=q_1,\ldots,\overline{X}_{t_k}=q_k]$$

then the $1 \times n$ matrix

$$M(t_1,\ldots,t_n;q_1,\ldots,q_k) = [M_j(\ ;\ )]$$

is the probability distribution of the states of X at time t_k, as conditioned by these observations of $\overline{X}$. In particular, since $\overline{X}_{t_k}=q_k$, it follows that $M_j(\ ;\ )$ is nonzero only if $f(j)=q_k$, where $\{M_j(\ ;\ ) \mid f(j)=q_k\}$ gives the probability distribution of states inside the lump $f^{-1}(q_k)$.

So as to translate distributions of X back up into distributions of $\overline{X}$, let us suppose further that $\overline{X}$ has m states ($m \le n$) and, relative to some specified ordering of the lumps, let Q_ℓ denote the ℓ^{th} lump, i.e., the collection of sets

$$\{Q_\ell \mid 1 \le \ell \le m\}$$

is the partition of Q induced by f. Accordingly, if $\pi=[\pi_1\pi_2\cdots\pi_n]$ is a probability distribution over the states of X, we let $\overline{\pi}$ denote the corresponding distribution that is,

$$\overline{\pi} = [\overline{\pi}_1\ \overline{\pi}_2\ \ldots\ \overline{\pi}_m]$$

where

$$\overline{\pi}_j = \sum_{i \varepsilon Q_j} \pi_i \quad (1 \le j \le m).$$

In terms of the above notation, weak lumability can then be characterized as follows.

<u>Theorem 1</u>: Let X be a time-homogeneous Markov process with transition function P and a fixed initial state probability distribution. Let $\overline{X}$ be a functional of X and, for $s \in T$, define

$$\Delta_s = \{M(t_1,\ldots,t_k;q_1,\ldots,q_k) \mid t_k = s\}.$$

Then $\overline{X}$ is a Markov process if and only if, for all s, $t \in T$ such that $s \leq t$ and for all $\pi, \pi' \in \Delta_s$, $\overline{\pi} = \overline{\pi'}$ implies $\overline{\pi P(t-s)} = \overline{\pi' P(t-s)}$.

The proof of Theorem 1 is essentially a generalized form of the argument used to prove its discrete-time analog (see [10], pp. 133-134, Theorem 6.4.1). To illustrate its application, let us suppose that the system in question is a multicomputer comprised of three identical computer modules. Suppose further that modules fail independently and that each fails permanently with a constant failure rate λ. Then we can take the base model X to be the Markov process depicted by the state-transition-rate diagram of Figure 1.

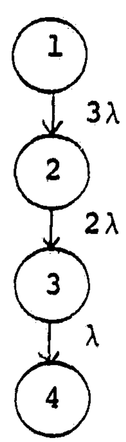

Figure 1

As for operational rates, let us assume they are normalized so that, at full capacity, the rate is 1 and with the loss of one or two modules, the rate is .5; loss of a third module results in total failure.

Accordingly, the operational structure in this case is the function

i	$f(i)$
1	1
2	.5
3	.5
4	0

and hence the functional $\overline{X}$ takes values in the state set $\overline{Q}=\{1,.5,0\}$. On taking the inverse of f, these states correspond as follows to lumps of Q:

$$1 \leftrightarrow \{1\}$$
$$.5 \leftrightarrow \{2,3\}$$
$$0 \leftrightarrow \{4\}.$$

If we now examine the probabilistic nature of $\overline{X}$, we find that the conditional probability $\Pr[\overline{X}_t=0|\overline{X}_s=.5]$ depends on the time that $\overline{X}$ enters state .5 from state 1 if the latter event is possible (i.e., if the probability of initially being in state 1 is non-zero). Thus, for example, if X is initially in state 1 with probability 1, i.e., $P=[1\ 0\ 0\ 0]$ is the initial state probability distribution, then we have such a dependence (on the past history of $\overline{X}$) and therefore $\overline{X}$ is not a Markov process. On the other hand, let us suppose the initial distribution is $P=[0\ 0\ 1\ 0]$, which is not a likely choice from a functional point of view but it serves to illustrate the role of p. In this case Δ_s (as defined in the statement of Theorem 1) is the same for any time s in T, i.e., it is the set

$$\Delta_s = \{[0\ 0\ 1\ 0],[0\ 0\ 0\ 1]\}.$$

Accordingly, the conditions of Theorem 1 are vacuously satisfied and, therefore, $\overline{X}$ is a Markov process for this choice of p. Moreover, it

should be obvious that $\overline{X}$, in this case, is time-homogeneous. Other distributions, such as p=[0 1 0 0] can be shown to result in Markov processes which are not time-homogeneous.

Turning our attention now to the second question (see Q2 above), the answer, in a slightly more specialized form, can be found in the existing literature (see [11], p. 1113-1114, Theorem 4). Stating the desired form of this result in terms of the notation defined above, we have:

<u>Theorem 2</u>: Let X be a time-homogeneous Markov process with generator matrix $A=[a_{ij}]$ and let $\overline{X}$ be a functional of X determined by f. Then $\overline{X}$ is a Markov process, whatever the initial distribution of X, if and only if for each $q \in \overline{Q}$ taken separately either

(i) For all $i,j \in Q$ such that $f(i) \neq q$ and $f(j)=q$,

$$a_{ij}=0$$

or

(ii) For all $r \in \overline{Q}$ such that $r \neq q$, the sum

$$\sum_{f(j)=r} a_{ij}$$

is the same for all $i \in Q$ such that $f(i)=q$.

Although the conditions of Theorem 2 guarantee that $\overline{X}$ is a Markov process relative to any initial distribution p for X, it should be noted that the specific nature of $\overline{X}$ (as specified by its transition function) will generally depend on p. Moreover, the process $\overline{X}$ need not be time-homogeneous.

To illustrate Theorem 2 and the above observations, let us again consider the Markov process X having the state-transition-rate diagram

given by Figure 1. Then, by (6), the generator matrix of X is the 4x4 matrix

$$A = \begin{vmatrix} -3\lambda & 3\lambda & 0 & 0 \\ 0 & -2\lambda & 2\lambda & 0 \\ 0 & 0 & -\lambda & \lambda \\ 0 & 0 & 0 & 0 \end{vmatrix} .$$

Suppose, however, that the operational sturcture in this case is one that corresponds to triplication with voting (TMR), i.e., the function

i	f(i)
1	1
2	1
3	0
4	0 .

Then $\overline{Q}=\{1,0\}$ and, applying Theorem 2, we see that state 1 (i.e., lump {1,2}) satisfies condition (i) and state 0 (i.e., lump {3,4}) satisfies condition (ii). Hence, the functional $\overline{X}$ is a Markov process. To determine the probabilistic nature of $\overline{X}$, let us rename state 0 (in $\overline{Q}$) as state 2 (permitting the use of standard matrix notation) and let $\overline{P}(s,t)$ denote the transition function of $\overline{X}$, i.e.,

$$\overline{P}(s,t) = [\overline{p}_{ij}(s,t)] \qquad (s\le t)$$

where

$$\overline{p}_{ij}(s,t) = \Pr[\overline{X}_t=j\,|\,\overline{X}_s=i].$$

Then, relative to an initial distribution $p=[p_1\ p_2\ p_3\ p_4]$ for X, if we let

$$d = \frac{p_1}{p_1+p_2}$$

it can be shown that the matrix $\overline{P}(s,t)$ has the following entries:

$$\left.\begin{aligned}
\overline{p}_{11}(s,t) &= \frac{e^{-2\lambda(t-s)}(1+2d(1-e^{-\lambda t}))}{(1+2d(1-e^{-\lambda s}))}\,,\\
\overline{p}_{12}(s,t) &= 1 - \overline{p}_{11}(s,t),\\
\overline{p}_{21}(s,t) &= 0,\\
\overline{p}_{22}(s,t) &= 1.
\end{aligned}\right\} \qquad (7)$$

From the above equations, we see that the transition function $\overline{P}(s,t)$ depends on d and, hence, on the initial distribution $p=[p_1\ p_2\ p_3\ p_4]$. Moreover, we observe that $\overline{X}$ is time-homogeneous (i.e., the values of $\overline{P}(s,t)$ depend only on the time difference t-s) just in case d=0. In other words, by the definition of d, $\overline{X}$ is time-homogeneous if and only if $p_1=0$, i.e., there is a zero probability that the underlying process X is initially in state 1 (all three modules fault-free). However, with our interpretation of $\overline{X}$ as a TMR model, this special case is pathological and hence, in cases of practical interest, $\overline{X}$ will not be time-homogeneous.

Finally, turning to the question of strong lumpability (see Q3 above), the answer can be characterized by removing condition (i) of Theorem 2 and modifying the proof to accommodate this change. More precisely, we have

<u>Theorem 3</u>: Let X be a time-homogeneous Markov process with generator matrix $A=[a_{ij}]$ and let $\overline{X}$ be a functional of X determined by f. Then $\overline{X}$ is a Markov process, whatever the initial distribution p of X and with a transition function that is independent of p, if and only if for each $q\in\overline{Q}$ the following condition is satisfied:

For all $r \in \bar{Q}$ such that $r \neq q$, the sum

$$\sum_{f(j)=r} a_{ij} \tag{8}$$

is the same for all $i \in Q$ such that $f(i)=q$.

To illustrate Theorem 3, suppose X is specified by the generator matrix

$$A = \begin{vmatrix} -3\lambda & \lambda & \lambda & \lambda & 0 & 0 & 0 \\ 0 & -2\lambda & 0 & 0 & \lambda & \lambda & 0 \\ 0 & 0 & -2\lambda & 0 & \lambda & 0 & \lambda \\ 0 & 0 & 0 & -2\lambda & 0 & \lambda & \lambda \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix} \tag{9}$$

and f is the function

q	f(q)
1	1
2	2
3	3
4	3
5	4
6	4
7	4

(10)

Testing condition (8) for states 1 and 2 in $\bar{Q}$, we see that it holds trivially since these states correspond to singleton lumps. As for state $3 \in \bar{Q}$ (corresponding to lump {3,4}), with respect to states $1,2 \in \bar{Q}$ the sums are zero for both i=3 and i=4; with respect to state $4 \in \bar{Q}$ the sum is 2λ for both i=3 and i=4. Thus condition (8) holds for

state 3. Finally, (8) is likewise satisfied for state $4 \in \bar{Q}$ and we conclude that $\bar{X}$ is a Markov process with a transition function that is independent of p.

In general, if X is strongly lumpable (as characterized by Theorem 3) it is easily shown that $\bar{X}$ must inherit the time-homogeneity of X. In other words, a strongly lumped process will always be time-homogeneous and, accordingly, it can be specified by a constant generator matrix. More precisely, if $\bar{Q}$ has m states and we rename them (if not already so named) with the integers from 1 to m, the generator matrix $\bar{A}=[\bar{a}_{qr}]$ of $\bar{X}$ can be constructed directly from A, where entry $\bar{a}_{qr}$ $(q \neq r)$ is given by the invariant sum of condition (8) for any i such that $f(i)=q$. (The diagonal entries $\bar{a}_{qr}$ are then determined by the condition that rows must sum to zero.) Thus, for the example just considered (see (9) and (10)), the generator matrix of $\bar{X}$ is the 4x4 matrix

$$A = \begin{vmatrix} -3\lambda & \lambda & 2\lambda & 0 \\ 0 & -2\lambda & 0 & 2\lambda \\ 0 & 0 & -2\lambda & 2\lambda \\ 0 & 0 & 0 & 0 \end{vmatrix} .$$

Reviewing the results of this section, it is clear that a strongly lumped process is the most desirable type of operational model. On the other hand, by Theorem 3, it is evident that such models require a relatively restricted "match" between the probabilistic nature of the base model (as specified by A) and the operational structure f.

The conditions of Theorem 2 are somewhat weaker although, when satisfied, the transition rates of the resulting Markov functional

are generally time-varying and dependent on the initial distribution of the underlying process. Of significance here is the fact that even without strong lumpability one can obtain operational models that are Markovian and admit to feasible, closed-form analytic solutions (see Equations (7), for example). What must be employed, in this case, are solution techniques for arbitrary (discrete-state) Markov processes, as opposed to more special (and much more familiar) techniques which apply only to time-homogeneous Markov processes.

Finally, regarding weak lumpability (Theorem 1), the requirement here is even less restrictive. However, depending on the nature of A, p and f, it may be very difficult to decide whether the condition of Theorem 1 is satisfied. Moreover, we currently know of no general means of solving such models without resorting to detailed computations at the base model level. The utility of weak lumpability is also curtailed by the fact that the initial state distribution of the base model is fixed. This may be satisfactory in certain applications but, in many cases, one wishes to examine the influence of different initial distributions. In such cases, one must derive a solution for each of the designated distributions provided, of course, that each admits to weak lumpability.

Theorems 1-3 thus provide formal support of what we, and others in the field, have observed through experience: at higher, more user-oriented levels of abstraction, it is difficult to maintain a Markovian representation of system behavior. As a consequence, we should seek means for accommodating operational models (functionals) that are not Markovian. The latter task is less formidable than it might appear if we bear in mind that, when evaluating a system S, an operational model $\overline{X}_S$ plays an intermediate role in support of a

designated performance variable Y_S. Thus, our knowledge of $\overline{X}_S$ can be restricted to that required to solve the probability distribution of Y_S, i.e., the performability of S. The latter observation serves as the guiding principle for the work described in the following section.

3. Solution Using Functionals

In this section, we consider the solution of performability with respect to a generally defined performance variable. This variable is defined in terms of an arbitrary operational model which, as discussed in the previous section, is generally non-Markovian. However, by relating this variable to the underlying Markov process, it is shown that system performabilities can still be solved using traditional Markov process methods.

3.1 Performance Variable

The performance variable we consider is motivated by needs which arise when performance is "recoverable" at the operational level. More precisely, let us say that an operational method $\bar{X} = \{\bar{X}_s \mid s \in [0,t]\}$ is recoverable if there is at least one operational rate q and time u,v $(0 \leq u < v \leq t)$ such that

$$Pr[\bar{X}_v > q \mid \bar{X}_u = q] > 0.$$

In other words, it is possible for the operational rate to recover from q to a value higher than q. If this is the case then, clearly, the operational rate at the end of utilization (i.e., the value $\bar{X}_t$) will generally not convey the "worst case" rate experienced during $[0,t]$. (On the other hand, if $\bar{X}$ is not recoverable, one can easily show that $\bar{X}_t$ is the worst case rate.)

Motivated by the above considerations, let us suppose that

$$X = \{X_s \mid 0 \leq s \leq t\}$$

is a finite-state Markov process and, relative to an operational structure f,

$$\bar{X} = \{f(X_s) \mid 0 \leq s \leq t\}$$

is the operational model (functional). Then a performance variable Y_t,

indicating the minimum operational rate during $[0,t]$, can be defined on $\mathcal{X}$ as follows:

$$Y_t = \min \{f(X_s) \mid 0 \leq s \leq t\} . \tag{11}$$

Figure 2 depicts the value of Y_t for a representative sample path of the functional $\mathcal{X}$. If $\mathcal{X}$ is <u>not</u> recoverable, then Y_t is simply the operational rate at time t. In this case, solving performability (relative to Y_t) is tantament to solving the transition function of $\mathcal{X}$. On the other hand, if $\mathcal{X}$ is recoverable, the solution process becomes more involved, as we discuss in the section that follows.

3.2 Performability Solution

As defined above, we note first that Y_t is a discrete performance variable since the base model X has a finite number of states and, hence, there is a finite number of operational rates. Therefore the performability p_S of S (see [4], definition 1a) is simply the probability distribution of Y_t, i.e.,

$$p_S(q) = \Pr[Y_t = q] . \tag{12}$$

Accordingly, if we are able to solve the conditional probabilities

$$g_{iq}(t) = \Pr[Y_t = q \mid X_0 = i]$$

where $i \in Q$, $q \in \overline{Q}$, $0 \leq t < \infty$, then p_S is obtained by the well-known formula

$$p_S(q) = \sum_{i \in Q} g_{iq}(t) \Pr[X_0 = i] .$$

Note that, in the case where $t = 0$, the minimum operational rate is just the operational rate associated with the initial state; in short

$$g_{iq}(0) = \begin{cases} 1 & \text{if } f(i) = q \\ 0 & \text{otherwise} . \end{cases} \tag{13}$$

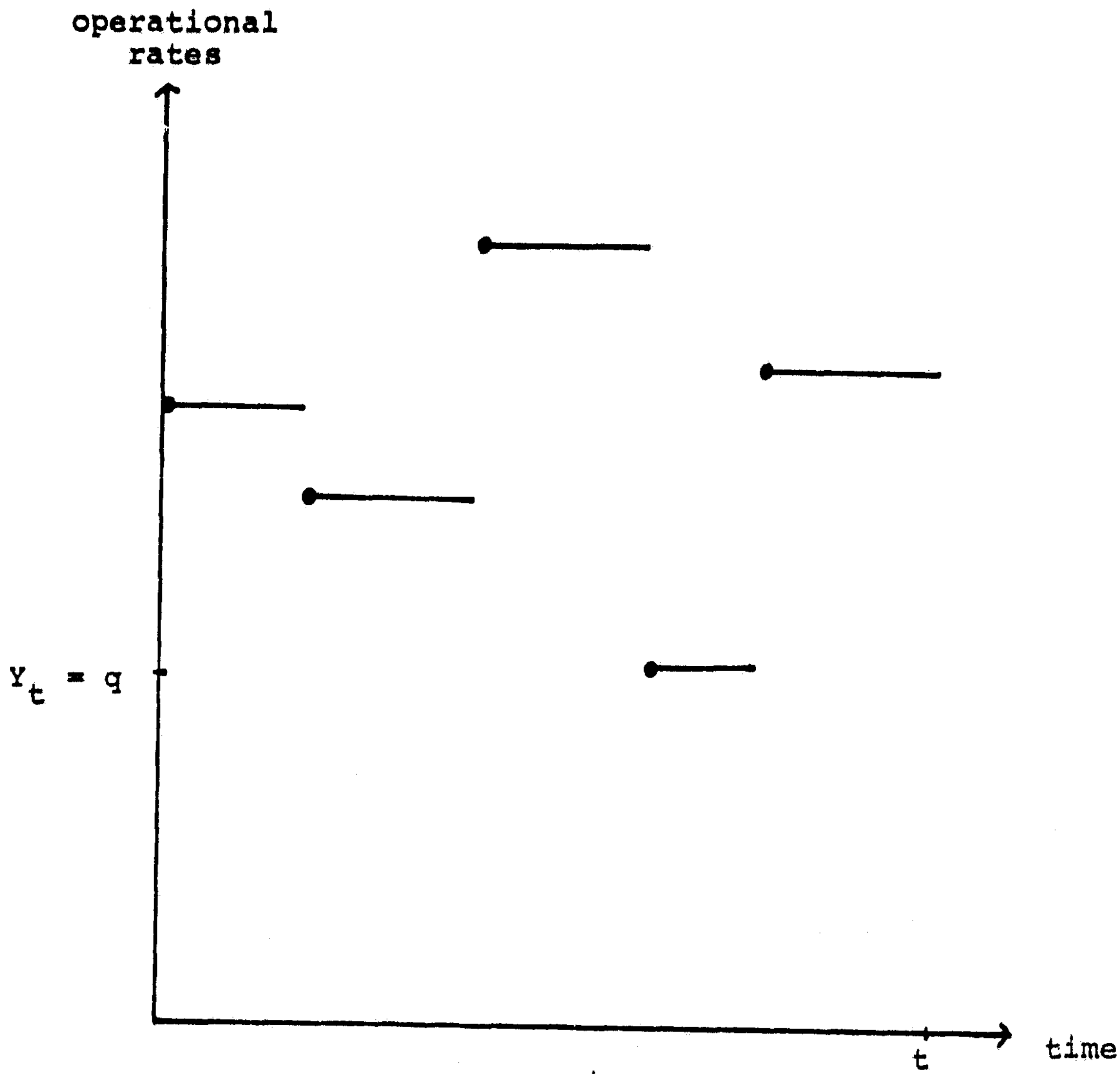

Figure 2

A sample path of an operational model and its corresponding minimum operational rate Y_t.

To relate the conditional probabilities $g_{iq}(t)$ to the generator matrix A of the underlying Markovian base model X (see (6)), we first introduce the random variable

$$\rho_i = \min \{s \mid s > 0,\ X_s \neq i \text{ and } X_0 = i\}$$

called the <u>first exit time</u> from state i (also called the <u>first sojourn time</u> in i). Then, using the "strong Markov property" at ρ_i (a detailed discussion of the strong Markov property can be found in [6], pp. 168-177), we are able to represent $g_{iq}(t)$ in terms of recurrence relations involving the coefficients a_{ij} of the generator matrix.

<u>Theorem 4</u>: Let X be a Markov process with an operational structure $f: Q \to \mathbb{R}$ then, for each $i \in Q$, $q \in \bar{\bar{Q}}$ and $0 \leq t < \infty$,

$$g_{iq}(t) = \begin{cases} \sum\limits_{f(j) \geq q} \int_0^t (1-\delta_{ij})\ a_{ij}\ e^{-a_{ii}\alpha} g_{jq}(t-\alpha)\,d\alpha & \text{if } f(i) > q \\[2ex] e^{-a_{ii}t} + \sum\limits_{r \geq q}\ \sum\limits_{f(j) \geq r} \int_0^t (1-\delta_{ij})\ a_{ij}\ e^{-a_{ii}\alpha} g_{jr}(t-\alpha)\,d\alpha & \text{if } f(i) = q \\[2ex] 0 & \text{otherwise.} \end{cases} \tag{14}$$

Proof: We first write

$$g_{iq}(t) = \Pr[Y_t = q,\ \rho_i > t \mid X_0 = i] + \Pr[Y_t = q,\ \rho_i \leq t \mid X_0 = i]. \tag{15}$$

Since $f(i) < q$ implies $g_{iq}(t) = 0$, we need to consider (15) only for two cases: $f(i) > q$ and $f(i) = q$.

Case 1: $f(i) > q$.

If the first state transition from state i occurs after t (i.e. $\rho_i > t$), then

$$Y_t = \min\{f(X_s) \mid 0 \leq s \leq t\} = f(i) \neq q.$$

It follows that

$$\Pr[Y_t = q,\ \rho_i > t \mid X_0 = i] = 0\ . \qquad (16)$$

On the other hand, if the event $\{\rho_i \leq t\}$ occurs, then it must be the case that the first state transition occurs at time ρ_i to some state j and the minimum operational rate occurs at a time between ρ_i and t. Then the strong Markov property at ρ_i gives

$$\Pr[Y_t = q \mid X_{\rho_i} = j\ ,\ X_0 = i] = \Pr[Y_{t-\rho_i} = q \mid X_0 = j].$$

Hence, applying the above equation,

$$\Pr[Y_t = q,\ \rho_i \leq t \mid X_0 = i]$$

$$= \int_0^t \Pr[Y_t = q,\ \rho_i = \alpha \mid X_0 = i]d\alpha$$

$$= \int_0^t \sum_{j \in Q} \Pr[Y_{t-\alpha} = q \mid X_0 = j]\ \Pr[X_\alpha = j,\ \rho_i = \alpha \mid X_0 = i]\ d\alpha$$

Now, since

$$\Pr[Y_{t-\alpha} = q \mid X_0 = j] = g_{jq}(t-\alpha)$$

and

$$\Pr[X_\alpha = j,\ \rho_i = \alpha \mid X_0 = i] = (1 - \delta_{ij})a_{ij}e^{-a_{ii}\alpha}\ ,$$

the above equality becomes

$$\Pr[Y_t = q,\ \rho_i \leq t \mid X_0 = i]$$

$$= \sum_{j \in Q} \int_0^t (1 - \delta_{ij})a_{ij}e^{-a_{ii}\alpha} g_{jq}(t-\alpha)d\alpha. \qquad (17)$$

Accordingly, replacing the right side of (15) by (16) and (17) and noticing that $g_{jq}(t-\alpha) = 0$ when $f(j) < q$, we have

$$g_{iq}(t) = \sum_{f(j)\geq q} \int_0^t (1 - \delta_{ij}) a_{ij} e^{-a_{ii}\alpha} g_{jq}(t-\alpha)\, d\alpha$$

Case 2: $f(i) = q$.

Since the sojourn times of a Markov process are exponentially distributed, we have

$$\begin{aligned} &\Pr[Y_t = q, \rho_i > t \mid X_0 = i] \\ &= \Pr[Y_t = q \mid \rho_i > t, X_0 = i] \Pr[\rho_i > t \mid X_0 = i] \\ &= 1 \cdot e^{-a_{ii}t} = e^{-a_{ii}t} . \end{aligned} \tag{18}$$

If the event $\{\rho_i \leq t\}$ occurs, then it must be the case that the first state transition occurs at time ρ_i to state j and $\min \{f(X_s) \mid \rho_i \leq s \leq t\} = r$ for some $r \geq q$. Accordingly, applying the strong Markov property at ρ_i and repeating the arguments for case 1, we have

$$\begin{aligned} &\Pr[Y_t = q, \rho_i \leq t \mid X_0 = i] \\ &= \sum_{r\geq q} \sum_{j\in Q} \int_0^t (1 - \delta_{ij}) a_{ij} e^{-a_{ii}\alpha} g_{jr}(t-\alpha)\, d\alpha . \end{aligned} \tag{19}$$

Replacing the right side of (15) by (18) and (19), we finally have, for the case that $f(i) = q$,

$$g_{iq}(t) = e^{-a_{ii}t} + \sum_{r\geq q} \sum_{f(j)\geq r} \int_0^t (1 - \delta_{ij}) a_{ij} e^{-a_{ii}\alpha} g_{jr}(t-\alpha)\, d\alpha .$$

To solve equation (14), we first note that the integrals on the rights side can be expressed as the convolutions of the functions $g_{jr}(t)$ and $W_{ij}(t) = (1 - \delta_{ij}) a_{ij} e^{-a_{ii}t}$, i.e.,

$$W_{ij} * g_{jr}(t) = \int_0^t W_{ij}(\alpha)\, g_{jr}(t-\alpha)\, d\alpha .$$

Then, since $W_{ij} * g_{jr}(t) = g_{jr} * W_{ij}(t)$, the derivative of the convolution $W_{ij} * g_{jr}(t)$ can be expressed as

$$\frac{d}{dt}\,[W_{ij}{*}\,g_{jr}(t)]$$

$$= \frac{d}{dt}\,[g_{jr}{*}\,W_{ij}(t)]$$

$$= \frac{d}{dt}\,\left[\int_0^t (1-\delta_{ij})a_{ij}e^{-a_{ii}(t-\alpha)}g_{jr}(\alpha)d\alpha\right]$$

$$= (-a_{ii})\int_0^t (1-\delta_{ij})a_{ij}e^{-a_{ii}(t-\alpha)}g_{jr}(\alpha)d\alpha$$

$$+ (1-\delta_{ij})a_{ij}g_{jr}(t)$$

$$= (-a_{ii})\ [g_{jr}{*}\,W_{ij}(t)] + (1-\delta_{ij})a_{ij}g_{jr}(t)$$

$$= (-a_{ii})\ [W_{ij}{*}\,g_{jr}(t)] + (1-\delta_{ij})a_{ij}g_{jr}(t).$$

Thus, taking derivative term by term and letting $f(i) > q$,

$$\frac{d}{dt}\,g_{iq}(t) = \sum_{f(j)\geq q} \frac{d}{dt}[W_{ij}{*}\,g_{jq}(t)]$$

$$= (-a_{ii})\ \left[\sum_{f(j)\geq q} W_{ij}{*}\,g_{jq}(t)\right]$$

$$+ \sum_{f(j)\geq q} [(1-\delta_{ij})a_{ij}g_{jq}(t)]$$

$$= (-a_{ii})\ g_{iq}(t) + \sum_{f(j)\geq q}[(1-\delta_{ij})a_{ij}g_{jq}(t)]$$

$$= a_{ii}g_{iq}(t) + \sum_{\substack{f(j)\geq q\\ j\neq i}} a_{ij}g_{jq}(t)$$

$$= \sum_{f(j)\geq q} a_{ij}g_{jq}(t). \tag{20}$$

Similarly, when $f(i) = q$,

$$\frac{d}{dt} g_{iq}(t) = (-a_{ii})e^{-a_{ii}t} + \sum_{r \geq q} \sum_{f(j) \geq r} \frac{d}{dt}[W_{ij} * g_{jr}(t)]$$

$$= (-a_{ii})e^{-a_{ii}t} + (-a_{ii}) \left[\sum_{r \geq q} \sum_{f(j) \geq r} W_{ij} * g_{jr}(t) \right]$$

$$+ \sum_{r \geq q} \sum_{f(j) \geq r} [(1 - \delta_{ij}) a_{ij} g_{jr}(t)]$$

$$= (-a_{ii})e^{-a_{ii}t} + (-a_{ii})[g_{iq}(t) - e^{-a_{ii}t}]$$

$$+ \sum_{r \geq q} \sum_{f(j) \geq r} [(1 - \delta_{ij}) a_{ij} g_{jr}(t)]$$

$$= (-a_{ii}) g_{iq}(t) + \sum_{r \geq q} \sum_{f(j) \geq r} [(1 - \delta_{ij}) a_{ij} g_{jr}(t)]$$

$$= \sum_{r \geq q} \sum_{f(j) \geq r} a_{ij} g_{jr}(t). \tag{21}$$

Accordingly, combining (20) and (21), we obtain the following expression for the derivative of $g_{iq}(t)$

Corollary: Let X be a Markov process with operational structure $f:Q \to \mathbb{R}$ then, for all $i \in Q$, $q \in \overline{Q}$ and $0 \leq t < \infty$,

$$\frac{d}{dt} g_{iq}(t) = \begin{cases} \sum_{f(j) \geq a} a_{ij} g_{jq}(t) & \text{if } f(i) > q \\ \sum_{r \geq q} \sum_{f(j) \geq r} a_{ij} g_{jb}(t) & \text{if } f(i) = q \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

The equation (22) is, in fact, a system of linear differential equations with $n \cdot m$ variables $g_{iq}(t)$ where $i \in Q$ and $q \in \overline{Q}$. Hence, by solving for $g_{iq}(t)$ in terms of the coefficients a_{ij} and time t,

the probability distribution of the random variable Y_t can be expressed as, for each $q\varepsilon\overline{Q}$,

$$\Pr[Y_t = q] = \sum_{i\varepsilon Q} g_{iq}(t) \Pr[X_0 = i].$$

3.3 Taboo Probability Functions

An alternative approach to derive expressions for the probability distribution of the random variable Y_t is to use the notion of "transition probability with taboo states" defined as

$$_{H}p_{ij}(t) = \Pr[X_s \notin H,\ 0 \leq s \leq t;\ X_t = j \mid X_0 = i] \tag{23}$$

where $i,j\varepsilon Q$, $H\subseteq Q$ and $t\geq 0$. In words, $_{H}p_{ij}(t)$ is the probability, starting from state i, of entering the state j at time t under the restriction that none of the states in the set H is entered during the period [0,t]. The set H is called the taboo set and $_{H}p_{ij}(t)$ is called the transition probability from i to j in time t under the taboo set H.

The analytic and probabilistic properties of taboo probability functions have been studied extensively and a detailed account is contained in [6]. (Note that the taboo transition probability defined above is the simplified form described in [6] on page 191.) To represent $_{H}p_{ij}(t)$ in terms of transition rates of the underlying Markov process, one can simply solve the transition function of the transformed Markov process obtained from the original underlying process by making states in H absorbing state (i.e., by deleting all the transitions from each state in H to other states in the state transition diagram of the original process). More rigorously, following arguments similar to the proofs of

Theorem 4 and its corollary, we can establish the validity of the above transformation method by the following Kolmogorov's differential equation:

<u>Theorem 5</u>: Let $X = \{X_t \mid t\varepsilon T\}$ be a continuous time finite state Markov process then, for all $i,j\varepsilon Q$ and $t \geq 0$,

$$\frac{d}{dt}\,{}_H p_{ij}(t) = \begin{cases} \sum\limits_{k\varepsilon Q} a_{ik}\,{}_H p_{kj}(t) & \text{if } i \notin H \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

To establish the relationship between the random variable Y_t and the taboo probabilities, we note that the complemented probability distribution function of Y_t can be expressed as

$$\begin{aligned} &\Pr[Y_t \geq q] \\ &= \Pr[\min\{f(X_s) \mid 0 \leq s \leq t\} \geq q] \\ &= \Pr[f(X_s) \geq q,\ 0 \leq s \leq t] \\ &= \Pr[X_s \notin \{k \mid f(k) < q\},\ 0 \leq s \leq t]. \end{aligned}$$

That is, $\Pr[Y_t \geq q]$ can be regarded as the taboo probability in time t under the taboo set $\{k \mid f(k) < q\}$. Hence if we let the generator matrix of the transformed process be a $n\times n$ dimensional matrix

$$\Gamma_q = [\gamma_{ij}] \tag{25}$$

such that for all i, j in Q,

$$\gamma_{ij} = \begin{cases} a_{ij} & \text{if } i\varepsilon \{k \mid f(k) \geq q\} \\ 0 & \text{otherwise,} \end{cases}$$

then by Theorem 5, for all $i,j\varepsilon Q$ and $t \geq 0$,

$$\frac{d}{dt} \Pr[Y_t \geq q,\ X_t = j \mid X_0 = i] = \sum_{k\varepsilon Q} \gamma_{ik} \Pr[Y_t \geq q,\ X_t = j \mid X_0 = k] \tag{26}$$

Solving the above system of differential equations, we obtain

$$\Pr[Y_t \geq q] = p(0)\, e^{\Gamma_q t}\, \mathbb{1}_q \qquad (27)$$

where p(0) is the initial probability distribution of the underlying process and $\mathbb{1}_q$ is a n x 1 matrix

$$\mathbb{1}_q = [c_i] \qquad (28)$$

such that

$$c_i = \begin{cases} 1 & \text{if } f(i) \geq q, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, since

$$\Pr[Y_t = q] = \Pr[Y_t \geq q] - \Pr[Y_t \geq q + 1],$$

the probability distribution of the random variable Y_t can be expressed as, for all $q \varepsilon \bar{Q}$,

$$\Pr[Y_t = q] \qquad (29)$$

$$= p(0)\, e^{\Gamma_q t}\, \mathbb{1}_q - p(0)\, e^{\Gamma_{q+1} t}\, \mathbb{1}_{q+1}$$

where p(0) is the initial probability distribution of the underlying Markov process X.

3.4 Conditional Expected Operational Rate

In contrast with the above approach of regarding the performance of a system to be the minimum operational rate experienced by the system during [0,t], the performance of the system can also be taken to be the "expected" operational rate of the system during the period. For instance, when the operational rates of a system can be associated with the rates of information processed in the unit time, then the expected operational rate can often be interpreted as expected throughput rate, expected capacity rate, expected

production rate, etc. The performance variable can be used to characterize the overall performance of both non-degradable and degradable computers (see [2] and [8], for example). However, in order to analyze the performance of a degradable computer in detail, we found it is useful to decompose the expected operational rate into components involving different operational modes of the system.

More precisely, if we denote the average operational rate of a system by the random variable

$$\delta_t = \frac{1}{t} \int_0^t f(X_s)ds \tag{30}$$

where [0,t] is the utilization period of the system, then the expected operational rate of the system can be decomposed according to the following well-known formula:

$$E[\delta_t] = \sum_{q\varepsilon\overline{Q}} E[\delta_t \mid Y_t = q] \Pr[Y_t = q] \tag{31}$$

where Y_t is given by (11). In other words, the effect on the system performance (relative to $E[\delta_t]$) of an operational mode having an associated operational rate q can be quantified by taking the product of the conditional expectation $E[\delta_t \mid Y_t = q]$ and the probability of occurance of the operational mode $\Pr[Y_t = q]$.

To solve the conditional expectations $E[\delta_t \mid Y_t = q]$, let us write P_i to represent the conditional probability given $X_0 = i$ and write $E_i[\delta_t \mid Y_t = q]$ instead of $E[\delta_t \mid Y_t = q, X_0 = i]$. Then, by expanding the conditional expectations, we have, for all $q\varepsilon\overline{Q}$,

$$E[\delta_t \mid Y_t = q] = \sum_{i\varepsilon Q} E_i[\delta_t \mid Y_t = q] \Pr[X_0 = i \mid Y_t = q]. \tag{32}$$

Moreover, for all $q \epsilon \bar{Q}$,

$$E_i[\delta_t \mid Y_t = q]$$

$$= \int_0^\infty P_i[\delta_t > \alpha \mid Y_t = q] d\alpha$$

$$= \int_0^\infty \frac{P_i[\delta_t > \alpha,\ Y_t = q]}{P_i[Y_t = q]} d\alpha$$

$$= \frac{P_i[Y_t \geq q]}{P_i[Y_t = q]} \int_0^\infty P_i[\delta_t > \alpha \mid Y_t \geq q] d\alpha$$

$$- \frac{P_i[Y_t \geq q+1]}{P_i[Y_t = q]} \int_0^\infty P_i[\delta_t > \alpha \mid Y_t \geq q+1] d\alpha$$

$$= \frac{P_i[Y_t \geq q]}{P_i[Y_t = q]} E_i[\delta_t \mid Y_t \geq q] - \frac{P_i[Y_t \geq q+1]}{P_i[Y_t = q]} E_i[\delta_t \mid Y_t \geq q+1] \quad (33)$$

Since $P_i[Y_t = q]$, $P_i[Y_t \geq q]$ and $P_i[Y_t \geq q+1]$ are given by (27) and (29), we only need to solve for $E_i[\delta_t \mid Y_t \geq q]$ and $E_i[\delta_t \mid Y_t \geq q+1]$. Now, because E_i is a linear operator, for all $q \epsilon \bar{Q}$,

$$E_i[\delta_t \mid Y_t \geq q]$$

$$= \frac{1}{t} \int_0^t E_i[f(X_s) \mid Y_t \geq q] ds$$

$$= \frac{1}{t} \int_0^t \sum_{j \epsilon Q} f(j)\ P_i[X_s = j \mid Y_t \geq q] ds$$

$$= \frac{1}{t\ P_i[Y_t \geq q]} \left\{ \sum_{j \epsilon Q} f(j) \int_0^t P_i[X_s = j,\ Y_t \geq q] ds \right\}. \quad (34)$$

To evaluate the integration, note that

$$P_i[X_s = j,\ Y_t \geq q]$$

$$= P_i[X_s = j;\ f(X_\alpha) \geq q \quad \text{for all } 0 \leq \alpha \leq t]$$

$$= P_i[f(X\ ) \geq q \text{ for all } s < \alpha \leq t | X_s = j;\ f(X_\alpha) \geq q \text{ for all } 0 \leq \alpha \leq s]\ P_i[X_s = j;\ f(X_\alpha) \geq q \text{ for all } 0 \leq \alpha \leq s].$$

Hence, by Markov properties,

$$P_i[X_s = j,\ Y_t \geq q]$$
$$= P_i[X_s = j,\ Y_s \geq q]\ P_j[Y_{t-s} \geq q]. \tag{35}$$

To simplify the notations, we observe that $P_i[Y_{t-s} \geq q]$and $P_i[X_s = j,\ Y_s \geq q]$ are the taboo probabilities under the taboo set $\{k|f(k) < q\}$. Thus, if we let

$${}_qP_{ij}(t) = \text{taboo transition probability from } i \text{ to } j \text{ in time } t \text{ under } \{k|f(k) < q\}$$

and

$${}_qD_i(t) = \sum_{j\varepsilon Q} {}_qP_{ij}(t) \ ,$$

then by combining (34) and (35)

$$E_i[\delta_t|Y_t \geq q]$$
$$= \frac{1}{t\ {}_qD_i(t)} \sum_{j\varepsilon Q} f(j)\ [{}_qD_j * {}_qP_{ij}(t)]$$

where ${}_qD_j * {}_qP_{ij}(t)$ is the convolution of ${}_qD_j(t)$ and ${}_qP_{ij}(t)$. Also by (33),

$$E_i[\delta_t|Y_t = q]$$
$$= \frac{1}{t\ P_i[Y_t = q]} \sum_{j\varepsilon Q} f(j) [{}_qD_j * {}_qP_{ij}(t)]$$
$$- \frac{1}{t\ P_i[Y_t = q]} \sum_{j\varepsilon Q} f(j) [{}_{q+1}D_j * {}_{q+1}P_{ij}(t)]. \tag{36}$$

Finally, by (33) and (36), the conditional expected operational rate of a given value of Y_t can be expressed as

$$E[\delta_t | Y_t = q]$$

$$= \frac{1}{t \Pr[Y_t = q]} \sum_{i,j \in Q} p_i(0)\ f(j)[{}_qD_j{}^*{}_qP_{ij}(t)]$$

$$- \frac{1}{t \Pr[Y_t = q]} \sum_{i,j \in Q} p_i(0)\ f(j)[{}_{q+1}D_j{}^*{}_{q+1}P_{ij}(t)]$$

where $p(0) = [p_1(0), \ldots p_n(0)]$ is the initial probability distribution vector of the underlying Markov process X.

3.5 An Example

The example considered here is essentially a Markov model of a TMR system where the simplex system has failure rate λ (see Figure 3). It is assumed that the system has coverge factor c as well as capability of software error recovery. In Figure 3, each state of the process is denoted by two numbers separated by a slash: i/q. The integer i represents the number of operational modules; a prime is appended (i') if the state denotes a software error recovery state. The number q, specifies a relative worth of the state, i.e., q is the operational rate of the state i. The transition rates from each state are noted on the state diagram. If the utilization period is $T = [0,t]$, then the system performance is given by

$$Y_t = \min \{f(X_s) | 0 \leq s \leq t\}.$$

By examining the state transition graph, it follows immediately that the generator matrix is

$$A = \begin{bmatrix} -(3\lambda+10^{-3}) & 3\lambda c & 10^{-3} & 3\lambda(1-c) \\ 0 & -2\lambda & 0 & 2\lambda \\ 10^3 & 0 & -(10^2+10^3) & 10^2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

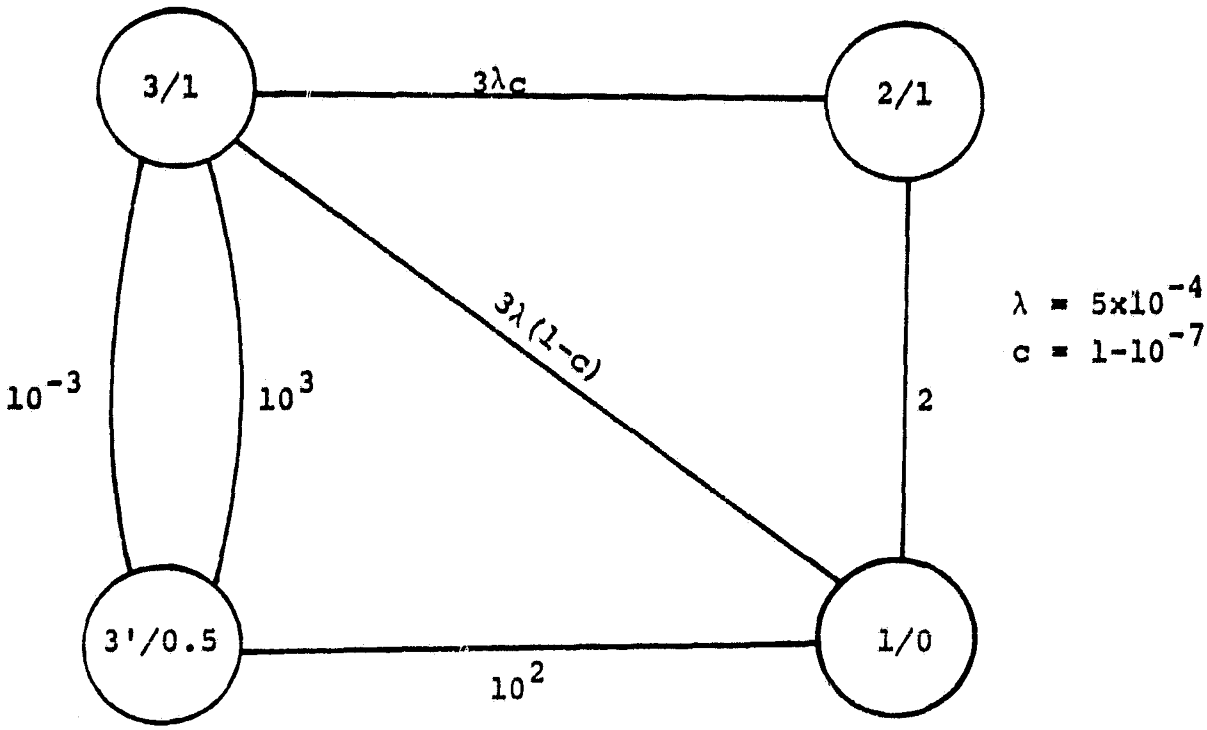

Figure 3

Markov model of a TMR system
with software error recovery

If we let

$$\Gamma_1 = \begin{bmatrix} -(3\lambda+10^{-3}) & 3\lambda c & 10^{-3} & 3\lambda(1-c) \\ 0 & -2\lambda & 0 & 2\lambda \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\Gamma_{0.5} = \begin{bmatrix} -(3\lambda+10^{-3}) & 3\lambda c & 10^{-3} & 3\lambda(1-c) \\ 0 & -2\lambda & 0 & 2\lambda \\ 10^3 & 0 & -(10^2+10^3) & 10^2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and $\Gamma_0 = A$, then by solving equation (24) the taboo transition probability functions are given by

$$_1P(t) = [_1p_{ij}(t)]$$

$$= \begin{bmatrix} e^{-0.0025t} & (1-10^{-7})(e^{-0.001t}-e^{-0.0025t}) & 0 & 0 \\ 0 & e^{-0.001t} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$_{0.5}P(t) = [_{0.5}p_{ij}(t)]$$

$$= \begin{bmatrix} c_1 & c_2 & c_3 & 0 \\ 0 & e^{-0.001t} & 0 & 0 \\ c_4 & c_5 & c_6 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

where

$$c_1 = 0.9999991735\ e^{-0.0015909t} - 0.0000008265\ e^{-1100.000909t}$$

$$c_2 = -2.538498242(e^{-0.0015909t}-e^{-0.001t})$$

$$c_3 = 0.0000009092(e^{-0.0015909t}-e^{-1100.000909t})$$

$$c_4 = 0.9090915(e^{-0.0015909t}-e^{-1100.000909t})$$

$$c_5 = -2.307727772(e^{-0.0015909t}-e^{-0.001t})$$

$$c_6 = 0.0000008265\ e^{-0.0015909t} + 0.9999991735\ e^{-1100.000909t}.$$

and

$$ {}_0P(t) = P(t) = e^{At}I $$

$$ = \begin{bmatrix} c_1 & c_2 & c_3 & 1-(c_1+c_2+c_3) \\ 0 & e^{-0.001t} & 0 & 1-e^{-0.001t} \\ c_4 & c_5 & c_6 & 1-(c_4+c_5+c_6) \\ 0 & 0 & 0 & 1 \end{bmatrix} $$

where c_i i= 1,2,...,6 are the same as in ${}_{0.5}P(t)$.

To determine the probability distribution of the random variable Y_t, notice that equation (29) can be rewritten as

$$ Pr[Y_t = q] = p(0)\ {}_qP(t)\ \mathbb{1}_q - p(0)\ {}_{q+1}P(t)\ \mathbb{1}_{q+1}. $$

Assuming that the system initially has all three modules operational, i.e.,

$$ p(0) = [1,0,0,0] \ , $$

then, by (12), the performability of the system relative to the performance variable Y_t can be expressed as

$$ p_S(1) = Pr[Y_t = 1] $$

$$ = (1-10^{-7})e^{-0.001t} + 10^{-7}e^{-0.0025t} $$

$$ p_S(0.5) = Pr[Y_t = 0.5] $$

$$ = -1.538498159\ e^{-0.0015909t} + 1.538498342\ e^{-0.001t} $$

$$ - 10^{-7}\ e^{-0.0025t} $$

$$ p_S(0) = Pr[Y_t = 0] $$

$$ = 1 + 1.538498159\ e^{-0.0015909t} - 2.58498242\ e^{-0.001t}. $$

In Figure 4 the performability of the system is represented by three curves: I, II and III. Curve I is the probability that the system will operate at operational level 1 throughout the utilization period. The probability is equal to 1 at time 0 and gradually diminishes to 0 as the length of T goes to infinity. Curve II is the probability that the system suffered performance degradation during the utilization period while remaining operational throughtout the period. As shown in the figure, the probability of degraded performance goes from 0 to some positive probability and decreases to 0 again when t increases. This is due to the fact that state 1 is an absorbing state, thus, the system will eventually end up in failure state. Finally, we note that curve III is the familiar S-shape function for system failure probability.

To compute the conditional expected throughput rate of the performance variable Y, we also assume that p(0) = [1,0,0,0]. Then,

$$E[\delta_t | Y_t = 1]$$

$$= \frac{1}{t \Pr[Y_t = 1]} \sum_{j \in Q} f(j) [{}_1D_j * {}_1P_{3j}(t)]$$

$$= \frac{1}{t \Pr[Y_t = 1]} [{}_1D_3 * {}_1P_{33}(t) + {}_1D_2 * {}_1P_{32}(t)]$$

$$= \frac{10^{-7} t\, e^{-0.0025t} + (1-10^{-7}) t\, e^{-0.001t}}{t \Pr[Y_t = 1]}$$

$$= 1.$$

Similarly, carrying out the computations, we have

$$E[\delta_t | Y_t = 0.5]$$

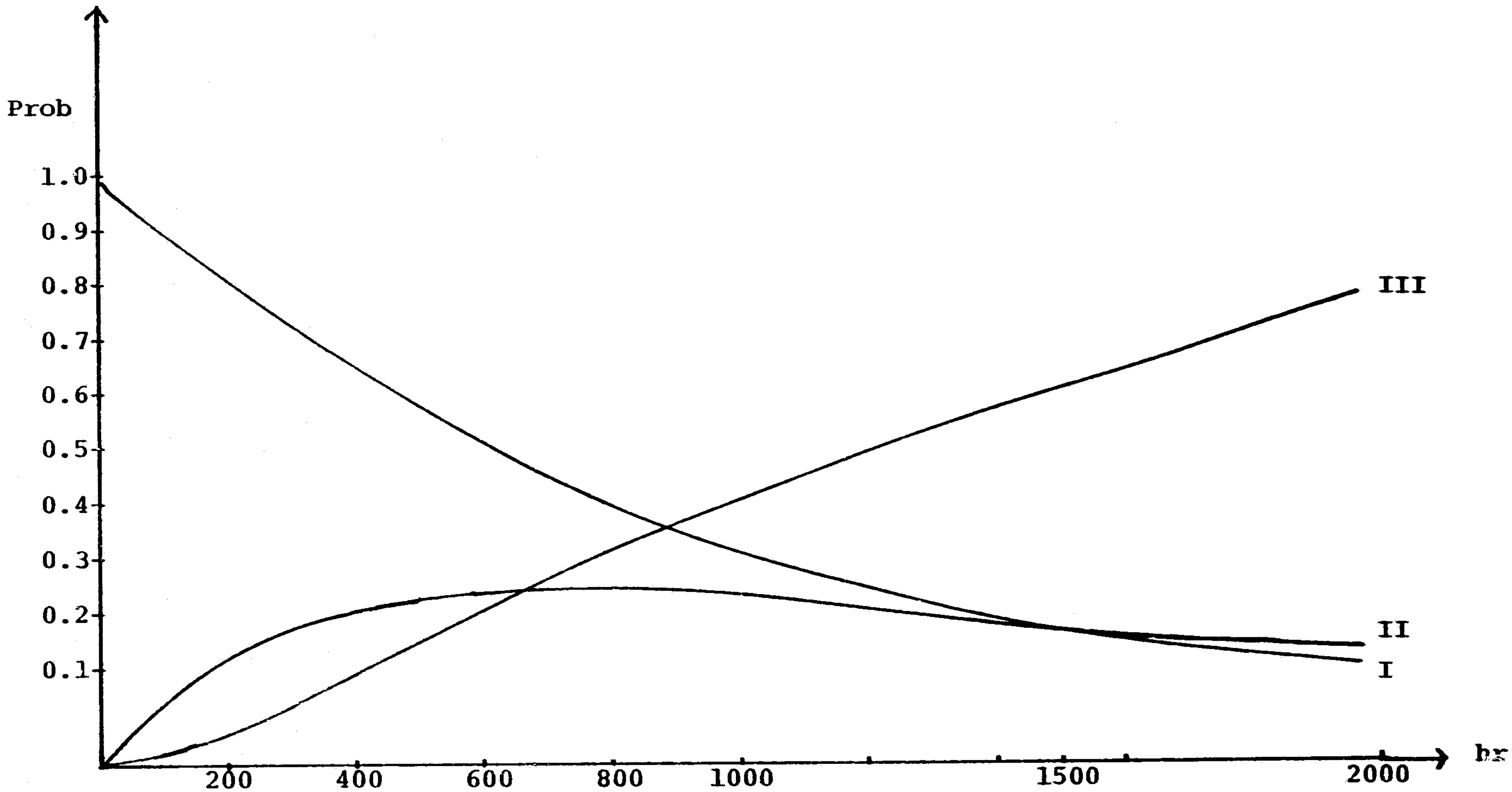

Figure 4

Performability of the system as a function of time

$$= \frac{1}{t \Pr[Y_t = 0.5]} \sum_{j \varepsilon Q} f(j)[{}_{0.5}D_j * {}_{0.5}P_{3j}(t)]$$

$$- \frac{1}{t \Pr[Y_t = 0.5]} \sum_{j \varepsilon Q} f(j)[{}_1D_j * {}_1P_{3j}(t)]$$

$$= \frac{1}{t \Pr[Y_t = 0.5]} (1.538498342 \ t \ e^{-0.001t}$$

$$- 1.538497523 \ t \ e^{-0.0015909t} + 0.0017755848 \ e^{-0.0015909t}$$

$$- 0.0017755848 \ e^{-0.001t} - 10^{-7} \ e^{-0.0025t})$$

and

$$E[\delta_t | Y_t = 0]$$

$$= \frac{1}{t \Pr[Y_t = 0]} [-964.0617977 \ (1-e^{-0.0015909t})$$

$$+ 2538.498242 \ (1-e^{-0.001t})$$

$$+ 1.538497523 \ t \ e^{-0.0015909t}$$

$$- 2.538498242 \ t \ e^{-0.001t}$$

$$- 0.0017755848 \ e^{-0.0015909t}$$

$$+ 0.0017755848 \ e^{-0.001t}].$$

Accordingly, if we assume that the utilization period of the system is [0,t] where t = 10 hours, then the performability and the corresponding conditional expected throughput rate are tabulated in Table 1.

$t = 10$ hrs	Performability ($Pr[Y_t = q]$)	Conditional Expected Operational Rate	$E[\delta_t \mid Y_t = q]\ Pr[Y_t = q]$
$q = 1$	.9900498323	1	.9900498323
$q = 0.5$	.0089740727	.9999641244	.0087737507
$q = 0$	.000976095	.511416195	.0004991908

Table 1

Conditional Expected Operational Rate

REFERENCES

[1] A. Costes, C. Landrault and J.-C. Laprie, "Reliability and availability models for maintained systems featuring hardware failures and design faults", IEEE Trans. Comput., Vol. C-27, pp. 548-560, June 1978.

[2] H. Kobayashi, Modeling and Analysis: An Introduction to System Performance Evaluation Methodology. Reading, MA: Addison-Wesley, 1978.

[3] R. E. Barlow and F. Proschan, Statistical Theory of Reliability and Life Testing Probability Models. New York, NY: Holt, Rinehart and Winston, 1975.

[4] J. F. Meyer, "On evaluating the performability of degradable computing systems", IEEE Trans. Comput., Vol. C-29, pp. 720-731, August 1980.

[5] J. F. Meyer, D. G. Furchtgott and L. T. Wu, "Performability evaluation of the SIFT computer", IEEE Trans. Comput., Vol. C-29, pp. 501-509, June 1980.

[6] K. L. Chung, Markov Chains. Berlin: Springer-Verlag, 1960.

[7] M. D. Beaudry, "Performance-related reliability measures for computing systems", IEEE Trans. Comput., Vol. C-27, pp. 540-547, June 1978.

[8] F. A. Gay and M. L. Ketelsen, "Performance evaluation for gracefully degrading systems", Proc. 1979 Int. Symp. Fault-Tolerant Computing, Madison, WI, June 1979, pp. 51-58.

[9] E. Cinlar, Introduction to Stochastic Processes. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[10] J. G. Kemeny and J. L. Snell, Finite Markov Chains. Princeton, NJ: Van Nostrand, 1960.

[11] C. J. Burke and M. Rosenblatt, "A Markov function of a Markov chain", Annals of Math. Statistics, Vol. 29, pp. 1112-1122, 1958.